Available online at www.sciencedirect.com

SCIENCE DIRECT

Virology 331 (2005) 325–337

VIROLOGY

www.elsevier.com/locate/yviro

Complete genomic nucleotide sequence and analysis of the temperate bacteriophage VWB

W. Van Dessel[a], L. Van Mellaert[a], H. Liesegang[b], C. Raasch[b], S. De Keersmaeker[a], N. Geukens[a], E. Lammertyn[a], W. Streit[b], J. Anné[a,*]

[a]*Laboratorium voor Bacteriologie, Katholieke Universiteit Leuven, Rega Instituut, Minderbroedersstraat 10, B-3000 Leuven, Belgium*
[b]*Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Grisebachstrasse 8, 37077 Göttingen, Germany*

Received 1 September 2004; returned to author for revision 21 September 2004; accepted 16 October 2004
Available online 14 November 2004

Abstract

The entire double-stranded DNA genome of the *Streptomyces venezuelae* bacteriophage VWB was sequenced and analyzed. Its size is 49,220 bp with an overall molar G + C content of 71.2 mol%. Sixty-one potential open reading frames were identified and annotated using several complementary bioinformatics tools. Clusters of functionally related putative genes were defined, supporting a refined version of the modular theory of phage evolution.

Keywords: VWB; Bacteriophage; Streptomyces; Genome sequence; Sequence analysis

Introduction

Bacteriophages can be detected in almost every biological niche and they are probably the most abundant type of organisms on earth (Paul et al., 2002; Pedulla et al., 2003). Therefore, phages are thought to play major roles in the ecological balance of microbial life and in microbial diversity.

In contrast to their abundance, relatively few phages have been isolated and investigated. Well-studied groups of phages are those infecting *Escherichia coli*, of which the best characterized is bacteriophage λ (Campbell, 1994), and those affecting lactic acid bacteria (Brüssow, 2001; Proux et al., 2002). The genomic sequence from about 200 phages has been published, and due to their relatively small genomes (30–300 kb) and the improvement of sequencing techniques, the number of sequenced phage genomes is increasing exponentially.

Although several phages infecting *Streptomyces* spp. have been identified (Anné et al., 1984; Lomovskaya et al., 1980; Smith, in press), the genome sequence of only two of these phages, ϕC31 and ϕBT1, both isolated using *Streptomyces coelicolor* as a host, has been reported (Gregory, 2003; Smith et al., 1999). Here, we present the third completed genomic sequence of a phage infecting *Streptomyces* spp. This temperate phage, bacteriophage VWB, belongs to the group of *Siphoviridae* (Anné et al., 1984) and has been shown to infect *Streptomyces venezuelae* ETH14630 and *Streptomyces exfoliatus* ATCC12672. In this study, we present an extensive analysis of the VWB genome, as well as evidence supporting a refined version of the modular theory of phage evolution.

* Corresponding author. Microbiology and Immunology, Laboratory of Bacteriology, K.U. Leuven, Rega Institute, Minderbroedersstraat 10, B-3000 Leuven, Belgium. Fax: +32 16 337340.
E-mail addresses: Wesley.Vandessel@pandora.be (W. Van Dessel), Lieve.Vanmellaert@rega.kuleuven.ac.be (L. Van Mellaert), hlieseg@gwdg.de (H. Liesegang), craasch@gwdg.de (C. Raasch), Sophie.Dekeersmaeker@rega.kuleuven.ac.be (S. De Keersmaeker), Nick.Geukens@rega.kuleuven.ac.be (N. Geukens), Elke.Lammertyn@rega.kuleuven.ac.be (E. Lammertyn), wstreit@gwdg.de (W. Streit), Jozef.Anne@rega.kuleuven.ac.be (J. Anné).

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2004.10.028

Results and discussion

Identification of putative open reading frames

The complete dsDNA genome of phage VWB (GenBank accession number NC_005345) has been sequenced as described in the materials and methods section. The obtained sequence consists of 49,220 bp and contains 71.2% G + C. The *cos* site (Fig. 1), characteristic for a non-headful packaging mechanism, was localized previously by restriction mapping (Anné et al., 1985). Since the exact *cos* sequence has not yet been determined, the numbering of nucleotides was initiated at the *attP* site. Putative open reading frames (ORFs) larger than 150 bp were predicted using Genemark and Glimmer, resulting in 61 potential ORFs (Table 1). Most of the ORFs, except genes 1, 18, 20 21, 46, 58, and 61, are transcribed from the left to the right (Fig. 1).

The overall molar G + C content of the coding parts of the phage genome is 71.2 mol%, which is consistent with the G + C content for known *Streptomyces* spp. (average 70 mol%) (Nakamura et al., 2000). Due to this high G + C level, codon usage in *Streptomyces* is biased. Generally, the amount of codons containing a 3rd letter G or C, instead of A or T, exceeds 90% in Streptomycetes. In silico support for the existence of the obtained phage ORFs was acquired from the calculation of this 3rd letter GC content of each ORF, using FramePlot 2.3 (Ishikawa and Hotta, 1999). Eight putative ORFs contained between 70% and 80% 3rd letter GC, whereas the level was between 80% and 90% in 21 other ORFs. In the 32 remaining ORFs, this value was above 90%. The total value of 88.5% 3rd letter GC in VWB ORFs is comparable to that of *Streptomyces* (Nakamura et al., 2000). Furthermore, 3rd letter GC levels which are more than 15% lower than those of the host have been reported for ORFs of *S. coelicolor* phages ϕC31 and ϕBT1. Taking these points into account, we think that the proposed ORFs are acceptable.

Linked to the biased codon use, the distribution of codons used in the phage ORFs was compared to that of 'the average ORF' from *S. venezuelae* (Nakamura et al., 2000) using the GCUA program. In 53 phage ORFs, this distribution was very similar to 'the average host ORF'. In eight ORFs, at least the predominantly used codon was conserved. In one case only (gene 59), no link concerning preferential codon use could be established between phage and host. These findings add further support to our ORF mapping.

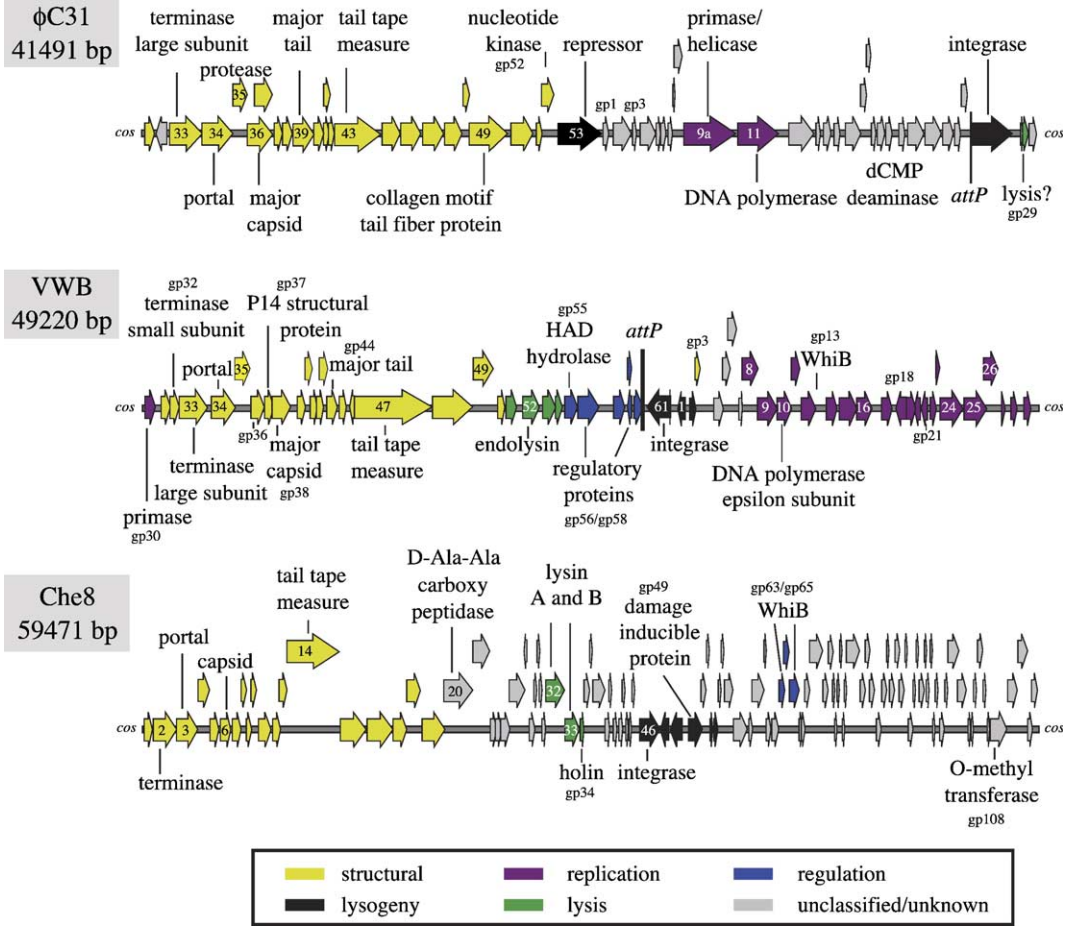

Fig. 1. Genome comparison of *Streptomyces* phage VWB, *Streptomyces* phage ϕC31 and mycobacteriophage Che8. Putative functions are marked by name. Annotation and numbering of Che8 and ϕC31genes was taken from Pedulla et al. (2003) and Smith et al. (1999), respectively.

Table 1
Overview of phage VWB ORFs and summary of homology searches

VWB ORFS				HOMOLOGY SEARCHES				
gp	Start	Stop	AA	Homology	Tool	Location (VWB)	Score	*E*-value
1	2314	1931	127	–				
2	2578	2892	104	–				
3	2870	3103	77	*Streptomyces coelicolor* A3(2) genome	blastn	nt 143-190	56	4.E-06
4	3884	4405	173	–				
5	4402	4668	88	–				
6	4665	5150	161	–				
7	5273	5485	70	–				
8	5482	6318	278	gp1 phages ϕBT1 and ϕC31 *Streptomyces coelicolor*	blastp	aa 1-278	51	1.E-05
9	6320	7399	359	gp3 phages ϕBT1 *Streptomyces coelicolor* A3(2)	blastp	aa 1-243	47	7.E-05
				gp3 phages ϕC31 *Streptomyces coelicolor* A3(2)	blastp	aa 1-243	46	2.E-04
10	7399	8190	263	*Corynebacterium efficiens* YS-314 DNA	blastn	nt 126-203	60	8.E-06
				putative DNA polymerase III, epsilon subunit *Streptomyces avermitilis* MA-4680	blastp	aa 1-263	143	2.E-33
				putative DNA polymerase III, epsilon subunit *Streptomyces coelicolor* A3(2)	blastp	aa 1-263	137	1.E-31
11	8183	8653	156	–				
12	7809	9485	258	PAS46 phage ϕAsp2 *Actinoplanes*	blastp	aa 83-258	60	2.E-08
13	9485	9730	81	putative WhiB-family transcriptional regulator *Streptomyces avermitilis* MA-4680	blastp	aa 1-81	76	2.E-08
				sporulation regulatory protein *Streptomyces coelicolor* A3(2)	blastp	aa 1-81	76	7.E-13
				gp68 mycobacteriophage Che9d	blastp	aa 1-81	61	2.E-08
				gp65 mycobacteriophage Che8	blastp	aa 1-81	61	2.E-08
				gp49 mycobacteriophage TM4	blastp	aa 1-81	57	2.E-07
14	10219	10662	147	–				
15	10829	11785	318	RNaseE *Xanthomonas campestris*	blastp	aa 1-318	50	8.E-06
16	11788	12588	266	ORF10 from oleandomycin biosynth. gene cluster *Streptomyces antibioticus*	blastn	nt 93-123	54	5.E-04
17	13167	13667	166	–				
18	14364	13939	141	–				
19	14368	14988	206	–				
20	15225	14989	78	hypothetical protein *Streptomyces avermitilis* MA-4680	blastp	aa 1-78	56	8.E-07
				SLV.8 from linear plasmid pSLV45 *Streptomyces lavendulae*	blastp	aa 1-78	54	2.E-06
				SLV.3 from linear plasmid pSLV45 *Streptomyces lavendulae*	blastp	aa 1-78	53	4.E-06
21	15599	15321	92	–				
22	15770	16102	110	–				
23	16099	16254	51	–				
24	16326	17642	438	neocarzinostatin biosynth. locus *Streptomyces neocarzinostaticus*	blastn	nt 161-508	174	3.E-40
				Streptomyces coelicolor A3(2) complete genome	blastn	nt 187-623	168	2.E-38
				plasmid SCP1 *Streptomyces coelicolor*	blastn	nt 199-623	145	3.E-31
				Streptomyces avermitilis MA-4680 complete genome	blastn	nt 167-623	131	4.E-27
				hypothetical membrane protein *Streptomyces avermitilis* MA-4680	blastp	aa 1-230	263	3.E-70
				probable integral membrane protein *Streptomyces avermitilis* MA-4680	blastp	aa 1-230	254	2.E-67

(continued on next page)

Table 1 (*continued*)

VWB ORFS				HOMOLOGY SEARCHES				
gp	Start	Stop	AA	Homology	Tool	Location (VWB)	Score	*E*-value
				putative integral membrane protein *Streptomyces coelicolor* A3(2)	blastp	aa 1-230	251	1.E-66
				putative membrane protein *Streptomyces coelicolor* A3(2)	blastp	aa 1-230	250	2.E-66
25	17668	18918	416	–				
26	18747	19508	253	hypothetical protein *Streptomyces avermitilis* MA-4680	blastp	aa 1-253	65	5.E-10
				hypothetical protein *Mycobacterium paratuberculosis*	blastp	aa 1-93	50	5.E-05
27	19682	19846	54	–				
28	20207	20491	94	–				
29	20950	21345	131	–				
30	21946	22527	193	putative primase *Staphylococcus aureus* MW2	blastp	aa 36-193	74	2.E-12
				ORF11 phage ϕ105 *Bacillus subtilis*	blastp	aa 36-193	66	5.E-10
				putative DNA-polymerase/DNA-primase phage ϕadh *Lactobacillus*	blastp	aa 36-193	58	1.E-07
				putative primase phage A2 *Lactobacillus casei*	blastp	aa 36-193	55	7.E-07
31	22849	23304	151	–				
32	23304	23783	159	gp1 mycobacteriophage Che8	blastp	aa 1-159	46	6.E-04
33	23862	25472	536	hypothetical protein *Streptomyces avermitilis* MA-4680	blastn	nt 323-431	66	3.E-07
				gp13 mycobacteriophage Bxz2	blastp	aa 1-536	214	3.E-55
				gp13 mycobacteriophage L5	blastp	aa 1-536	212	8.E-55
				gp13 mycobacteriophage D29	blastp	aa 1-536	206	1.E-52
				gp10 mycobacteriophage Bxb1	blastp	aa 1-536	192	8.E-49
				gp2 mycobacteriophage Che8	blastp	aa 1-536	113	8.E-25
				gp33 *Streptomyces coelicolor* phage ϕBT1	blastp	aa 1-536	92	2.E-18
				gp33 *Streptomyces coelicolor* phage ϕC31	blastp	aa 1-536	89	1.E-17
34	25621	26925	434	gp14 mycobacteriophage D29	blastp	aa 1-434	83	1.E-15
				gp14 mycobacteriophage L5	blastp	aa 1-434	62	1.E-09
				gp3 mycobacteriophage Che8	blastp	aa 1-434	78	2.E-14
				gp11 mycobacteriophage Bxb1	blastp	aa 1-434	67	2.E-11
				putative portal gp5 mycobacteriophage TM4	blastp	aa 1-434	66	1.E-10
				gp14 mycobacteriophage Bxz1	blastp	aa 1-434	61	3.E-09
35	26918	27721	267	gp4 mycobacteriophage Che8	blastp	aa 1-267	159	2.E-38
36	27778	28815	245	ori region of plasmid pIJ101 *Streptomyces lividans*	blastn	nt 153-183/584-595	62	2.E-06
37	28534	28923	129	structural protein P14 phage VWB *Streptomyces venezuelae* ETH14630	experimental			
38	28938	29984	348	structural protein P38 phage VWB *Streptomyces venezuelae* ETH14630	experimental			
				hypothetical phage protein *Streptococcus pyogenes* M18	blastp	aa 35-340	73	1.E-12
				ORF31 phage O1205 *Streptococcus thermophilus*	blastp	aa 35-340	57	6.E-08
				put. major head protein prophage Lj965 *Lactobacillus johnsonii*	blastp	aa 35-340	55	3.E-07
39	30342	30752	136	–				
40	30740	31102	120	–				
41	31107	31379	90	–				
42	31382	31777	131	–				
43	31576	31971	131	–				
44	31971	32525	184	structural protein P20 phage VWB *Streptomyces venezuelae* ETH14630	experimental			
				structural protein prophage λSa1 *Streptococcus agalactiae*	blastp	aa 1-184	74	2.E-12

Table 1 (*continued*)

VWB ORFS				HOMOLOGY SEARCHES				
gp	Start	Stop	AA	Homology	Tool	Location (VWB)	Score	*E*-value
				hypothetical protein phage SM1 *Streptococcus mitis*	blastp	aa 1-184	72	6.E-12
				major tail protein phage LC3 *Lactococcus lactis*	blastp	aa 1-184	72	9.E-12
				gp14 mycobacteriophage Che9d	blastp	aa 1-184	71	1.E-11
				structural protein phage rlt *Lactococcus lactis*	blastp	aa 1-184	70	4.E-11
45	32638	32988	116	–				
46	33422	33255	55	–				
47	33423	37745	1440	*Streptomyces coelicolor* A3(2)	blastn	nt 2892-2978	86	7.E-13
				Streptomyces avermitilis MA-4680	blastn	nt 2892-2991	64	3.E-06
				gp43 phage ϕBT1 *Streptomyces coelicolor* A3(2)	blastp	aa 22-485	258	3.E-69
				gp43 phage ϕC31 *Streptomyces coelicolor* A3(2)	blastp	aa 22-485	241	3.E-64
				tail component prophage Pi1 *Lactococcus lactis*	blastp	aa 1-1140	112	3.E-25
				putative minor capsid corynephage BFK20	blastp	aa 1-1140	106	1.E-23
				putative tape measure protein phage A118 *Lysteria monocytogenes*	blastp	aa 1-1140	100	7.E-22
				tail length tape measure protein phage ϕ13 *Streptomyces aureus*	blastp	aa 1-1140	98	5.E-21
48	37757	39994	745	–				
49	39961	41082	373	PAS30 phage ϕAsp2 *Actinoplanes*	blastp	aa 1-373	83	8.E-16
50	41330	41776	148	FluMu defective tail fiber potein Mu-like prophage *Haemophilus influenzae*	blastp	aa 1-148	54	2.E-06
51	41791	42384	197	–				
52	42725	43633	302	endolysin phage μ1/6 *Streptomyces aureofaciens*	blastn	nt 28-56/298-351	52	2.E-03
				endolysin phage μ1/6 *Streptomyces aureofaciens*	blastp	aa 1-192	205	2.E-52
53	43801	44448	215	–				
54	44493	44801	102	–				
55	45049	45693	214	put. hydrolase (HaloAcid Dehalogenase) *Methanococcus jannaschii*	blastp	aa 1-214	51	1.E-05
56	45737	46918	393	putative transcriptional regulator *Streptomyces avermitilis* MA-4680	blastp	aa 1-393	80	2.E-15
57	47697	48365	222	–				
58	48679	48449	76	putative regulatory protein *Streptomyces avermitilis* MA-4680	blastp	aa 1-76	51	3.E-06
				putative regulatory protein *Streptomyces coelicolor* A3(2)	blastp	aa 1-76	50	7.E-06
59	48485	48703	72	–				
60	48807	49214	135	–				
61	1495	212	427	integrase phage VWB *Streptomyces venezuelae* ETH14630	experimental			
				putative integrase *Streptomyces avermitilis* MA-4680	blastp	aa 1-427	108	3.E-25
				integrase bacteriophage pMLP1 *Micromonospora carbonaceae*	blastp	aa 1-427	93	3.E-21

VWB ORFs are arranged according to their position (start-stop) in the genome. Significant database matches are given in the column marked 'homology'. Tools used to search for similarity are blastn (nucleotide Blast search) or blastp (protein Blast search). Location of the region of homology in the VWB gene or gp is noted in the column 'location'. Scores and *E* values obtained in the Blast searches, are given in the last two columns.

Annotation of the predicted VWB ORFs

In order to assign functions to the gene products emerging from the predicted ORFs, databases at the European Molecular Biology Laboratory [EMBL Nucleotide Sequence database, release 80 and Uniprot, release 2.6] were interrogated for genes homologous to the putative phage ORFs. Searches were performed at nucleotide level, as well as at amino acid (aa)

level, using Blast 2 (Altschul et al., 1997) and Fasta 3 (Pearson and Lipman, 1988). Hits yielding scores >50 and corresponding E values ≤ 1.0 × E-3 were considered significant. The most significant hits are summarized in Table 1.

Very few hits (8) were counted at the nucleotide level, even with parameters set to very low stringency, and most of them are related to hypothetical genes in the complete genomes of *Streptomyces avermitilis* and *S. coelicolor*. This is no new issue, since it has been reported for several other phages (Lavigne et al., 2003; Pedulla et al., 2003). Also parallel with reported findings are the results for searches at the aa level. Adjustment of the search parameters to moderate stringency returned similarities to database proteins for 24 of the VWB gene products (gps). Experimental evidence is available for 4 gps (gp38, 44 and 61 and gp37) (Anné et al., 1995; Van Mellaert et al., 1998), although no significant database match was found for gp37. Sixteen of the 61 VWB gps show homology to those of known phages. The discovered homology allowed the assignment of a putative function in only nine of the 16 cases. Additionally, six of the 61 VWB gps are similar to bacterial proteins of which a (putative) function is known, resulting in a predicted function to 16 of the 61 VWB gps. Identity scores ranged from 21% to 65%. Alternatively, gps were examined for characteristic protein motifs using Pfam as an indication towards their putative function (Bateman et al., 2004). A specific motif is likely to be present in several ORFs, as will be discussed throughout the next section of this report.

Overview of functional gene clusters

Only to a minority of the putative ORFs, a putative function could be assigned. Nevertheless, the VWB genome can be split into five provisional clusters of genes with related functions. One cluster is supposed to support DNA replication, the others encode structural, regulatory, lytic, and lysogenic functions (Fig. 1). The defined clusters will be discussed in more detail.

Structural gene cluster

The cluster that could be defined best is the one containing the genes that encode the structural parts of the phage particle and the proteins needed in the assembly process. It embraces at least 19 ORFs (gene 32–50).

Most intriguingly is the relationship of gp32–35 to gp1–4 from mycobacteriophage Che8. VWB gp32–34 all show 20–25% identity to their respective Che8 counterparts, whereas this percentage is even higher for gp35. Although the conservation of the relative gene order in both phage genomes is remarkable, no clear explanation can be given for the observed difference in identity. Phage VWB thus might have acquired this 4-gene cluster from Che8 or a relative during one or more recombinatorial events in the distant past. Data concerning evolution are discussed in a separate section of this report.

An in-depth look at gp32 and 33 using Pfam reveals one motif in each gp, weakly similar to those of the small and the large subunits of a phage terminase, respectively. The same function was also predicted for gp32 from *Streptomyces* phage ϕC31, another homologous counterpart of VWB gp33. Exploration of the aa sequence of VWB gp34 did not bring in a significant domain match. However, this protein is 20% identical to the putative portal protein from mycobacteriophage TM4. Furthermore, the terminase and portal genes are apparently adjacent in phages of different origin, such as *Streptomyces* phage ϕC31, many of the mycobacteriophages (Pedulla et al., 2003), several *Streptococcus thermophilus* phages (Desiere et al., 1999), and also in lambdoid phages such as HK022 and HK97 (Juhala et al., 2000). Therefore, we postulate a similar situation in the VWB genome. Despite the rather high identity of VWB gp35 to gp4 of phage Che8 (40%) (Pedulla et al., 2003), no further indications of a potential function were found.

Gene homology searches on VWB gene 36 uncovered a small, highly conserved zone of 30 bp (nucleotide 153–183 of the ORF), similar to a 30 bp part of the *ori-rep* region of *Streptomyces lividans* plasmid pIJ101 (Kendall and Cohen, 1988), as well as to the 3′ terminal part of *rep* genes from other *Streptomyces* plasmids. Gp36 and several proteins of different origin, although no replication genes, all match this zone at the aa level. This may indicate a conserved functionality, but no motif confirming this was found.

Gp37, 38, and 44 have been identified experimentally as structural proteins by SDS-PAGE electrophoresis of the VWB virion proteins and N-terminal sequencing of the three most abundant proteins (Anné et al., 1995).

No significant database similarities to gp37 were found. Interestingly, however, two weak matches did show up: ORF7T and 7C from *Vibrio parahaemolyticus* phages VP16T and C, respectively. The corresponding genes are located near the putative capsid gene in the *Vibrio* phages. This organization is comparable to that in the VWB genome. Contrary to the situation of gp37, gp38 does match structural (head) proteins of several phages (Table 1). VWB gp38 and these proteins are approximately 20% identical. According to Desiere et al. (2000), a multiple alignment can be made between the sequences of the VWB gp38 major capsid protein and those of prophage Lj965, *S. thermophilus* phage Sfi11 and phage λ. This resulted in a detectable level of sequence conservation, as was confirmed by our analysis.

Next to gp37 and 38, gp44 was shown the third major structural protein (Anné et al., 1995). Current analysis using software based calculation of the protein size, however, suggests a size of 20 kDa, instead of 24 kDa as suggested by Anné et al. (1995).

Gp44 is significantly similar to structural (tail) proteins of several phages (Table 1). These proteins are 27–29% identical to VWB gp44. Alignment of these proteins shows 21 conserved amino acids. Most of these can be grouped in two major blocks (Fig. 2), but in those no functional motif was found. Surprisingly, the conserved blocks seem to be

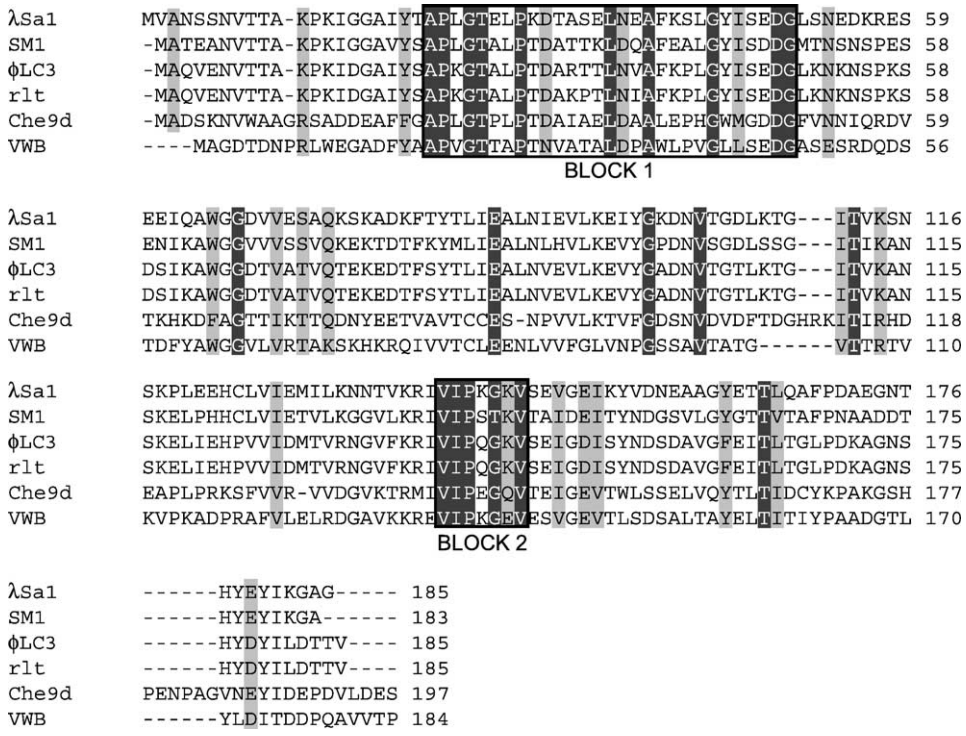

Fig. 2. Alignment of structural tail proteins from *Streptococcus agalactiae* prophage λSa1, *Streptococcus mitis* phage SM1, *Lactococcus lactis* phages LC3 and rlt, mycobacteriophage Che9d and *Streptomyces venezuelae* phage VWB. Conserved amino acids are indicated in dark grey, whereas residues with the same nature are indicated in light grey. Concentrated zones of conserved amino acids are boxed.

linked strictly to the aligned group of tail proteins, since they could not be found in tail proteins outside this group, not even in those from phages sharing the same host group.

The next elucidated ORF from the structural cluster, VWB gene 47, encodes a putative tail tape measure protein. The N-terminal half (aa 22–485) of this rather large protein (1441 aa residues) is quite conserved (32% identity) compared to gp43 from phage ϕC31 and ϕBT1, whereas the C-terminal half of gp47 contains only 6% maintained residues. Upon alignment of gp47 to its other database matches [several tail components, including tail tape measure proteins (TMP)], the observed similarity was lower, but more evenly distributed. The correlation between tail length and TMP size, according to Pedulla et al. (2003), also fits VWB gp47. Its 4323-bp gene length correlates well to its tail length (208.3 nm), as calculated from electron micrographs (Anné et al., 1984). The entire protein was searched for known patterns and domains, revealing the following results. First of all, no TMP repeat motif [WXXh, where X can be any residue and h is a hydrophobic residue (Pfam accession number PF05017)], as seen in the TMP from several *Lactococcus* and *Streptococcus* phages, could be detected in gp47. Nevertheless, other TMPs lacking this motif were identified previously, for instance, in several mycobacteriophages (Pedulla et al., 2003). Secondly, in the C-terminal part, a highly conserved transglycosylase SLT domain was found (aa 936–1060, Pfam accession number PF01464). A transglycosylase domain has been demonstrated before in at least two other phages. The TMP of mycobacteriophage Barnyard (Pedulla et al., 2003) contains a transglycosylase domain, albeit one of a slightly different type. Furthermore, an SLT domain was found in bacteriophage T7 gp16, a protein of approximately the same size as VWB gp47. It was shown nonessential for phage growth, but a beneficial effect was seen during infection of *E. coli* cells grown to high cell density conditions, in which murein is more highly cross-linked. In the absence of this SLT domain, internalization of the T7 phage genome was significantly delayed during infection (Moak and Molineux, 2000). A TMP with a similar function may be expected for VWB gp47, since the residues necessary for catalytic activity were quite conserved in both T7 gp16 and VWB gp47, as well as in *E. coli* SltY (Moak and Molineux, 2000) as shown in Fig. 3. Also, the four transmembrane helices predicted in gp47 might support this function. Indeed, several membrane-bound transglycosylases have been reported before in *E. coli* (Walderich and Holtje, 1991), thus adding further circumstantial evidence to this hypothesis.

Another gp from this structural cluster, gp49, was found homologous to PAS30 from *Actinoplanes* phage ϕAsp2 (Jarling et al., 2004). It was expected to be the major structural head protein from ϕAsp2. Although PAS30 itself does not show any homology to database gps, this assumption was deduced from the finding that PAS30 together with PAS27, which shows resemblance to tail proteins from other phages, are the most abundant proteins in the virion (Jarling et al., 2004). In VWB, gp38 has already been labeled the major virion protein (Anné et al.,

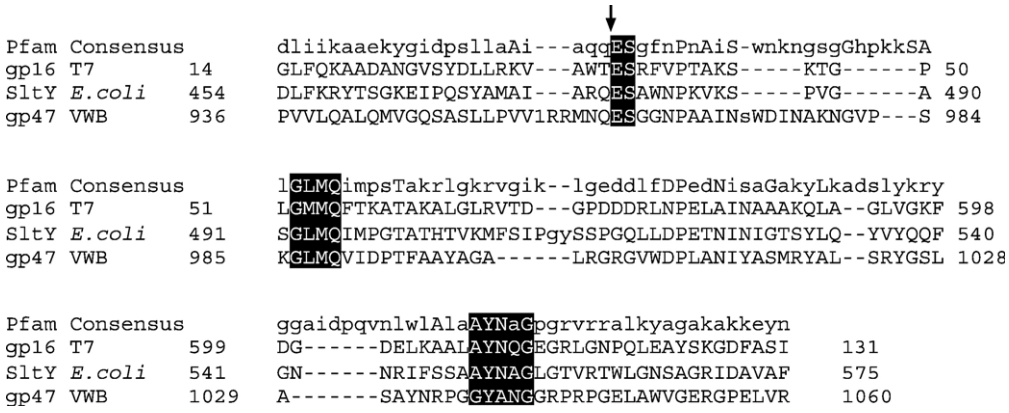

Fig. 3. Alignment of the transglycosylase domains of gp16 (phage T7), SltY (*E. coli*), gp47 (phage VWB) and the consensus motif used by Pfam for database searches. Grey boxes mark the regions required for catalytic activity. The catalytic glutamate is indicated by an arrow. Conserved residues are indicated by a capital in the Pfam consensus, others are shown in small caps.

1995), whereas gp49 is not one of the three most abundant proteins of VWB. Therefore, we postulate that gp49 may be a structural protein, but not the major head protein.

Gene 50, the final ORF within the VWB structural cluster showing significant homology to the database, is 26% identical to a defective tail fiber protein from a Mu-like prophage from *Haemophilus influenzae*. Although not excluding the option, no specific tail fiber repeat patterns (Pfam access numbers PF03335 and PF03406) were found. Consequently, several hypotheses on the nature of gp50 are still possible. It might be a minor part of the VWB tail structure, or alternatively, since electron micrographs of VWB do not show any tail fibers (Anné et al., 1984), this ORF might contain the remains of a damaged or discarded VWB tail fiber gene.

Without regard to the exact number of interspersed genes and the transcriptional direction of VWB gp46, the order of the terminase subunits (gp32–33), the major capsid component (gp38), the major tail protein (gp44), and the TMP (gp47) in VWB is similar to the genes in the late clusters of other *Siphoviridae* (Lucchini et al., 1998). Except for the terminase, the structural gene cluster of VWB resembles rather closely that of mycobacteriophage L5, and to a lesser extent, this seems also true for λ. Further research is required to determine the exact degree of similarity.

Lytic gene cluster

This cluster is very small, consistent to those encoding lysis genes in other phages, and is composed of 1 to a maximum of 4 ORFs (gene 52 + 51, 53, 54). Since database searches only offer a clue to the function of gp52, the border of the lytic gene cluster is not clearly defined. A homologue of the putative phage μl/6 endolysin from *Streptomyces aurofaciens* is found in gp52. Interestingly, the N-terminal part (aa 1–192) is well-conserved (68% aa identity), whereas the rest of the gp is only 26% identical, a fact shared, although less pronounced at the nucleotide level, since two small zones of approximately 50 bp in the DNA region corresponding to the N-terminal part, are 93% and 87% conserved, respectively. Despite this similarity, no reference to a motif suggesting a role in peptidoglycan catabolism was found in gp52 using Pfam searches. In contrast, a peptidoglycan binding domain is present in the C-terminal part of the μl/6 endolysin. This may suggest a yet unknown mechanism of lysis, or just a 'nonsense recombinatorial event' which happened between phages in the past.

Regulatory gene cluster

The boundaries of the remaining regulatory, lysogenic, and replicatory gene clusters are far less easy to define, since only a minority of the remaining ORFs allow the prediction of a function, based on database material. Therefore, borders might shift when new data become available in the future. Downstream the lytic gene cluster, the regulatory cluster stretches from genes 53 or 54 to gene 60. Nevertheless, one other putative regulatory gene, gene 13, was detected outside the confined regulatory region. This gene, one of the smallest genes in the VWB genome, is located in the replicatory cluster. Though the encoded protein counts only 83 aa, interrogations of the databases yielded very significant hits to WhiB-family-like transcriptional regulators from *S. coelicolor* and *S. avermitilis* (58% identity) and other sources, such as mycobacteriophages Che8, Che9d and TM4 (40–45% identity). As expected from these matches, a very strong WhiB domain was detected in gp13 (aa 10–73).

One of the proteins located within the actual regulatory cluster, VWB gp55, is most probably a member of the haloacid dehalogenase-like hydrolase family (HAD). It shares 25% identical aa with putative HAD hydrolases from *Methanococcus jannaschii* and *S. avermitilis*, as well as with several other hydrolases, although the resemblance to the latter is more limited. Furthermore, a highly probable HAD domain was found in gp55 (aa 2–181). The HAD family is known to catalyse diverse physiological functions in primary and secondary metabolism, membrane transport,

signal transduction, and nucleic acid repair. The vast majority of the known catalytic activities in the HAD family are directed at phosphoryl transfer, most of them ATPases and phosphatases (Allen and Dunaway-Mariano, 2004). More 'unusual' HAD family members, include those found in *Xanthomonas* and *Pseudomonas*. These allow growth on halogenated organic compounds as a sole carbon source. Furthermore, the polynucleotide kinase of phage T4 was also shown to be a member of this family. This enzyme is known to modify the ends of nicked tRNA, generated by a bacterial response to infection and facilitates repair by T4 RNA ligase (Galburt et al., 2002). Therefore, VWB gp55 may have either a regulatory function, a repair function or allows different host growth conditions. Further research is required to elucidate the correct function.

Next in line are gp56 and gp58. Both proteins match putative transcriptional regulators from *S. avermitilis* and *S. coelicolor*, and also from various other sources. The highly probable HTH-3 motif, which was detected in aa 12–66 and 20–74 from gp56 and 58, respectively, suggests a DNA-binding capability.

Lysogenic gene cluster

This cluster comprises rather few genes (probably gene 61–gene 2). Although the function of most of the gps is highly speculative, gp61 has experimentally been proven a site-specific integrase, belonging to the group of tyrosine recombinases. It matches quite a lot of integrases from different sources (Van Mellaert et al., 1998). Also, the attachment site allowing integration into the host genome is located in the region between gene 60 and 61 (bp 1–85). Based on the topology of the genes in the lysogenic clusters of lambdoid phages, a putative phage transcriptional repressor, maintaining the same transcriptional direction as the integrase gene, may be assumed near to gene 61. A possible candidate would be gp1. Unfortunately, no strong homologous repressor gene could be distilled from database interrogations. Gp1 and also gp2 are only insignificantly related to some regulatory genes. Nevertheless, a weak helix–turn-helix motif, found in both protein sequences, might allow binding of these proteins to DNA, and thus a transcription repressing function could be suggested. Experimental evidence (unpublished results), however, does not support any of the reported phage repressing mechanisms as such. Further investigation is needed to reveal the true mechanism.

Replicatory gene cluster

The transition zone between the regulatory and the replicatory cluster is not sharply defined. We believe that the replicatory cluster contains at least 22 ORFs (gene 8 to gene 30). Unfortunately, no data allowing a probable annotation are available for genes 2–7, although one can interpret results from Pfam on gp3 as weak evidence for the presence of a Nu1-domain from aa 1 to 38. Nu1 is a phage λ DNA-packaging protein containing a low affinity ATPase, which is stimulated by the presence of non-specific DNA (Hwang and Feiss, 1999). If so, gene 3 is the second stray gene (next to gene 13) in the VWB genome.

The replicatory cluster also shows evolution-related data. Gp8 is significantly homologous to gp1 of *S. coelicolor* phages ϕBT1 and ϕC31, whereas gp9 is borderline significantly homologous to gp3 of those phages. Alignment of VWB gp9 to its counterparts showed an abrupt stop of similarity at aa 243 from the VWB gp, notwithstanding a predicted size of 359 aa. This observation might point to a recombinatorial event between VWB and ϕC31, or a relative in the past.

Another important result pops up when the gp10 sequence is blasted. Gene 10 possibly encodes a DNA polymerase ε subunit, highly similar to those from different sources. The ε subunit is responsible for the 3′–5′ exonuclease activity of the polymerase. This is also supported by the prediction of a highly probable exonuclease domain at aa 7–184 using Pfam. Since VWB does not provide genetic material for any other part of a DNA polymerase, this subunit might have been picked up during transition through a host. Furthermore, the ε subunit is not involved in promoter recognition, so there is no urging reason to believe that a new type of hetero-oligomeric DNA polymerase, built from phage and host monomers, may result in some kind of phage-directed transcription regulation.

Searches in this cluster, also revealed another interesting finding. Gp11, significantly homologous to PAS46 from *Actinoplanes* phage ϕAsp2 (Jarling et al., 2004), contains a weak RuvC domain (VWB gp11 aa 87–228), as does PAS46. These domains are typical in crossover junction endodeoxyribonucleases. Jarling et al. (2004) proposed that PAS46 might therefore play a role in the resolution of Holliday junctions during the life cycle of the phage. A similar function may be suggested for VWB gp11.

Rather unexpectedly, the C-terminal part of gp15 shows homology to *Xanthomonas* RNaseE. This may suggest the conservation of a domain shared by both proteins, although no known motif was available from Pfam. The protein is also very weakly similar to gp80, a protein encoded by the replicatory cluster of mycobacteriophage CJW1 (Pedulla et al., 2003). These findings add at least some circumstantial evidence to the replicatory nature of the predicted ORF.

When gp16 is explored with Pfam, a very weak zinc finger motif from prokaryotic topoisomerases (Ahumada and Tse-Dinh, 1998; Tse-Dinh and Beran-Steed, 1988) emerges at aa 176–186. This may indicate a potential topoisomerase-like function of this protein. Furthermore, a weak motif can be noticed from phenazine biosynthesis proteins A and B, which have implications on virulence, competition, and biological control in mammals (Mavrodi et al., 1998).

Interestingly, the same type of topoisomerase-like zinc finger motif is also present in gp18, and in this protein, the motif is only weakly preserved as well.

The next ORF of interest, gene 21, shows no detectable homology, but a weak HTH motif was found in its protein sequence, providing it with a DNA-binding capability.

Gp24 is similar to hypothetical membrane proteins from *Streptomyces* spp. Conservation can be seen both at nucleotide and aa level. Intriguingly, at the aa level, the N-terminus of VWB gp24 (aa 1–230) is highly conserved (57% aa identity). Four transmembrane helices can be detected in this zone, suggesting a localization at the cell membrane. The C-terminus, not containing any known motifs, shows homology to proteins from scattered origin and is far less conserved.

Transmembrane helices were also found in gp25 and 26 (4 and 1 helices, respectively), thus indicating a potential capacity to integrate into the cell membrane. No homologous proteins suggesting a specific function were found, but surprisingly, many of the gp25 matching hits were situated between aa 2 and 246. A weak GLUG motif was found in gp25 aa 294–325. This motif is seen in IgA1 peptidases, attached to the cell wall peptidoglycan by an amide bond. IgA1 protease selectively cleaves human IgA1 and is likely to be a pathogenicity factor in some pathogens, such as *Streptococcus pneumoniae* (Chiavolini et al., 2003). VWB gp26 on the other hand, matches hypothetical protein Q82C24 from *S. avermitilis*. The latter is considered to be encoded by a prophage, as can be deducted from the information on the website (http://avermitilis.ls.kitasato-u.ac.jp/), accompanying the report of Ikeda et al. (2003). This gene is located 18 genes upstream of a putative integrase/recombinase (Q82C06) and nine genes downstream of a putative large terminase subunit (Q82C33).

Finally, a putative primase, approximately 25% identical to several phage primases, was observed in gp30. These, however, all turned out much larger (±800 aa) than VWB gp30 (193 aa). Gp30 is significantly similar to the N-terminal 200 aa of these primases. Increasing the size of gp30 by ignoring the stop codon does not extend the length of the conserved zone. Also, no primase-specific motif can be detected. Therefore, this gene may also be the outcome of a nonsense recombination event.

Evidence to the modular theory of phage evolution

From the previous results, we can state that some parts of the VWB genome are related to mycobacterial and *Streptomyces* phages, which are also GC-rich phages. There is no indication, however, that a more or less stable degree of genetic similarity to one or more specific phages or organisms can be drawn through large parts of the phage VWB genome. On the contrary, ORF(s) related to a specific genome are grouped in clusters of one or more genes. The ORFs within a cluster, such as the above described VWB gene 32–35 and gene 8–9 clusters, generally share the same level of identity to (a) related organism(s). The clusters themselves are alternated with each other and with parts of the genome that seem to be unrelated to any specific organism. Furthermore, the borders of these clusters do not necessarily coincide with those of the clusters of related functionality. Based on similar observations in other phage genomes (Desiere et al., 1998; Pedulla et al., 2003), a new refined version of the modular evolution model, originally proposed by Susskind and Botstein (1978) is emerging. This adapted model postulates that illegitimate recombination takes place quasi-randomly along recombining phage genomes, resulting in a random mix of recombinant phage offspring. This also implicates that this process is not confined to gene borders. Many of the recombinant phages will be defective for phage growth as a consequence of misplaced recombination and will as such be eliminated. Those that result in viable phages give rise to a new type of phage population (Hendrix, 2002; Juhala et al., 2000). Apart from the clustering mentioned above, observations made during analysis of the VWB genome add some evidence to this random recombination model. Although homology is lost at the nucleotide level, proof of a recombination event still resides at the aa level of some VWB proteins. Indeed, a clear difference in residue conservation level between two parts of the protein can be seen in at least 2 hypothetical VWB proteins. Compared to their respective database-recovered counterparts, the N-terminal parts of gp47 and 52 are moderately to strongly conserved, whereas the C-terminal parts are far less identical, and both parts are divided by a clear border. No specific patterns, such as direct repeats or palindromes, were found in the border region of gp47 and 52. The absence of a general 'recombination pattern' seems to support random recombination events, by which not only one or more entire genes, but also parts of genes are exchangeable.

Since phage DNA is not separated from bacterial, plasmid, or other genetic elements present in the host, genetic exchange may also occur between them. As a consequence, phages significantly influence evolution, physiology, and pathogenicity of their hosts (Boyd et al., 2001; Dobrindt and Reidl, 2000; Wagner and Waldor, 2002). The opinion of Sonea and Paniset (1976) is even more extreme. They postulated that activities of plasmids, phages, and other DNA exchange devices make all the planet's bacteria into a single global superorganism. These events of lateral gene transfer are challenging the established phylogenetic tree of evolution. Based on these and several other observations, the replacement of the current tree by a 'reticulate net', in which all organisms are interconnected, is proposed (Doolittle, 1999; Lawrence et al., 2002; Rohwer and Edwards, 2002). If this theory is correct, the lack of similarity between the phage ORFs and the database genes at the DNA level and the generally modest identity at the aa level must be explained by: (a) a lack of sequenced homologous genes in the database and (b) the accumulation of deletions/insertions and point mutations within a gene after the recombination event, which would lead to a diversification of genes, as was proposed by Desiere et al. (1998) during a study of *S. thermophilus*

bacteriophage Sfi21. The hypothesis concerning diversification by accumulation of deletions can be illustrated by observations made for VWB gp30 (primase). Indeed, the length of VWB gp30 is equal to the N-terminal part of its database homologues. The border region in the corresponding gene 30 contains an 11-bp direct repeat 6 bp after the stop codon (also followed by a 14- and 10-bp perfect palindromic repeat). Though short, this repeat might indicate a successful deletion event, which is potentially caused by 'slipping' of the DNA polymerase, a suggestion made in the mentioned study from Desiere et al. (1998) to explain a deletion event. The lack of significant nucleotide homology and the moderate level of aa identity to database material furthermore allow the presumption that this gene must have been received long ago. Although precaution is required, a timetable of relative gene acquirement may thus be composed for each phage or organism.

Materials and methods

Phage propagation and isolation of VWB genomic DNA

Phage VWB was propagated on *S. venezuelae* ETH14630. *S. venezuelae* was cultivated on phage medium (Korn et al., 1978). High titre lysates were prepared as described by Anné et al. (1984). DNA was isolated from the obtained phages according to a procedure described by Kieser et al. (2000). The obtained DNA was subcloned into pBR322 as described by Anné et al. (1985).

DNA sequencing

A minimal set of the overlapping clones described by Anné et al. (1985), was selected and partially digested with *Sau*3A. Fragments of 0.5–3.5 kb were isolated after electrophoretic separation on agarose gels, cloned into pTZ19R (Amersham Biosciences, Essex, UK) and sequenced using standard primers. Sequencing was performed using dye terminator technology on a model 377 sequencer (Applied Biosystems, Foster City, CA) or on Mega-BACE 1000/4000 capillary sequencers from Amersham Biosciences. A mean sequence coverage of 11.1-fold redundancy was obtained. The sequence is expected to be 99.999% accurate, or a statistical error of less than 1 in 10,000 bp. The GC-Phrap software package (http://www.phrap.org) was used to assemble the sequences. Editing and finishing was facilitated by the Staden software package (Staden et al., 2000). Sequencing of PCR-generated fragments was used to close single-stranded and double-stranded gaps.

Open reading frame analysis and annotation

Open reading frames (ORFs) were initially identified by the programs Glimmer (Delcher et al., 1999) and GeneMark (Besemer and Borodovsky, 1999; Besemer et al., 2001; Lukashin and Borodovsky, 1998). Third letter GC levels were verified using FramePlot 2.3 (Ishikawa and Hotta, 1999). The cut-off limit for ORFs without database homologues was 150 bp. Predicted ORFs and intergenic regions were used to interrogate non-redundant protein databases with Blast2 programs via the BLAST website (http://www.ncbi.nlm.nih.gov/blast). ORFs were entered into the ERGO Integrated Genomics Bio-informatics Suite for genome annotation and metabolic reconstruction (Integrated Genomics, Chicago, USA). The predicted ORFs were subjected to two rounds of annotation (one automatic and one manual). Manual annotation was performed using Blast2 (nucleotide and protein level) and Fasta (nucleotide and protein level) (Pearson and Lipman, 1988), Pfam (Bateman et al., 2004) and the ERGO Bioinformatics Suite. Databases used were EMBL Nucleotide Sequence database release 80, Uniprot release 2.6 and Pfam release 15.0. The codon usage of VWB and *S. venezuelae* was correlated by the GCUA program (McInerney, 1998).

Nucleotide sequence accession number

The complete nucleotide sequence of phage VWB was deposited in the GenBank under accession number NC_005345.

Acknowledgments

This research was supported by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). W.V.D. and S.D.K. are IWT research fellows. N.G. is a postdoctoral fellow of K.U.Leuven Onderzoeksfonds (PDM/03/270). The authors are grateful for G. Gottschalk for helpful discussions concerning the sequencing and annotation of the phage sequences.

References

Ahumada, A., Tse-Dinh, Y.C., 1998. The Zn(II) binding motifs of *E. coli* DNA topoisomerase I is part of a high-affinity DNA binding domain. Biochem. Biophys. Res. Commun. 251, 509–514.

Allen, K.N., Dunaway-Mariano, D., 2004. Phosphoryl group transfer: evolution of a catalytic scaffold. Trends. Biochem. Sci. 29, 495–503.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Anné, J., Wohlleben, W., Burkardt, H.J., Springer, R., Pühler, A., 1984. Morphological and molecular characterization of several actinophages isolated from soil which lyse *Streptomyces cattleya* or *S. venezuelae*. J. Gen. Microbiol. 130, 2639–2649.

Anné, J., Verheyen, P., Volckaert, G., Eyssen, H., 1985. A restriction endonuclease map of *Streptomyces* phage VWB. Mol. Gen. Genet. 200, 506–507.

Anné, J., Fiten, P., Van Mellaert, L., Joris, B., Opdenakker, G., Eyssen, H.,

1995. Analysis of the open reading frames of the main capsid proteins of actinophage VWB. Arch. Virol. 140, 1033–1047.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L., Studholme, D.J., Yeats, C., Eddy, S.R., 2004. The Pfam protein families database. Nucleic Acids Res. 32, D138–D141 (Database issue).

Besemer, J., Borodovsky, M., 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. 27, 3911–3920.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.

Boyd, E.F., Davis, B.M., Hochhut, B., 2001. Bacteriophage–bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol. 9, 137–144.

Brüssow, H., 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55, 283–303.

Campbell, A., 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. 48, 193–222.

Chiavolini, D., Memmi, G., Maggi, T., Iannelli, F., Pozzi, G., Oggioni, M.R., 2003. The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. BMC Microbiol. 3, 14.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with Glimmer. Nucleic Acids Res. 27, 4636–4641.

Desiere, F., Lucchini, S., Brüssow, H., 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. Virology 241, 345–356.

Desiere, F., Lucchini, S., Brüssow, H., 1999. Comparative analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260, 244–253.

Desiere, F., Pridmore, R.D., Brüssow, H., 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275, 294–305.

Dobrindt, U., Reidl, J., 2000. Pathogenicity islands and phage conversion: evolutionary aspects of bacterial pathogenesis. Int. J. Med. Microbiol. 290, 519–527.

Doolittle, W.F., 1999. Phylogenetic classification and the universal tree. Science 284, 2124–2129.

Galburt, E.A., Pelletier, J., Wilson, G., Stoddard, B.L., 2002. Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. Structure (Camb.) 10, 1249–1260.

Gregory, M.A., 2003. Direct Submission to GenBank (accession number NC_004664), University of Nottingham, QMC, Nottingham, NG7 2UH, United Kingdom

Hendrix, R.W., 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61, 471–480.

Hwang, Y., Feiss, M., 1999. A mutation correcting the DNA interaction defects of a mutant phage lambda terminase, gpNu1 K35A terminase. Virology 265, 196–205.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Omura, S., 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. 21, 526–531.

Ishikawa, J., Hotta, K., 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. FEMS Microbiol. Lett. 174, 251–253.

Jarling, M., Bartkowiak, K., Pape, H., Meinhardt, F., 2004. The genome of phiAsp2, an *Actinoplanes* infecting phage. Virus Genes 29, 117–129.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299, 27–51.

Kendall, K.J., Cohen, S.N., 1988. Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. J. Bacteriol. 170, 4634–4651.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A. (Eds.), Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich.

Korn, F., Weingärtner, B., Kutzner, H.J., 1978. A study of twenty actinophages: morphology, serological relationship and host range. In: Freerksen, E., Tarnok, I., Thumin, J.H. (Eds.), Genetics of the Actinomycetales. Fisher G, Stuttgart, pp. 251–270.

Lavigne, R., Burkal'tseva, M.V., Robben, J., Sykilinda, N.N., Kurochkina, L.P., Grymonprez, B., Jonckx, B., Krylov, V.N., Mesyanzhinov, V.V., Volckaert, G., 2003. The genome of bacteriophage phiKMV, a T7-like virus infecting *Pseudomonas aeruginosa*. Virology 312, 49–59.

Lawrence, J.G., Hatfull, G.F., Hendrix, R.W., 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184, 4891–4905.

Lomovskaya, N.D., Chater, K.F., Mkrtumian, M., 1980. Genetics and molecular biology of *Streptomyces* bacteriophages. Microbiol. Rev. 44, 206–229.

Lucchini, S., Desiere, F., Brüssow, H., 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology 246, 63–73.

Lukashin, A.V., Borodovsky, M., 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26, 1107–1115.

Mavrodi, D.V., Ksenzenko, V.N., Bonsall, R.F., Cook, R.J., Boronin, A.M., Thomashow, L.S., 1998. A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2–79. J. Bacteriol. 180, 2541–2548.

McInerney, J.O., 1998. GCUA (General Codon Usage Analysis). Bioinformatics 14, 372–373.

Moak, M., Molineux, I.J., 2000. Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol. Microbiol. 37, 345–355.

Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28, 292.

Paul, J.H., Sullivan, M.B., Segall, A.M., Rohwer, F., 2002. Marine phage genomics. Comp. Biochem. Physiol., B Biochem. Mol. Biol. 133, 463–476.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U.S.A. 85, 2444–2448.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs Jr., W.R., Hendrix, R.W., Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F., Brüssow, H., 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184, 6026–6036.

Rohwer, F., Edwards, R., 2002. The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.

Smith, M.C.M., 2004. Molecular genetics of *Streptomyces* phages. In: Calendar, R. (Ed.), The Bacteriophages. Oxford Univ. Press, New York. In press.

Smith, M.C., Burns, R.N., Wilson, S.E., Gregory, M.A., 1999. The complete genome sequence of the *Streptomyces* temperate phage straight phiC31: evolutionary relationships to other viruses. Nucleic Acids Res. 27, 2145–2155.

Sonea, S., Paniset, M., 1976. Towards a new bacteriology. Rev. Can. Biol. 35, 103–167.

Staden, R., Beal, K.F., Bonfield, J.K., 2000. The Staden package, 1998. Methods Mol. Biol. 132, 115–130.

Susskind, M.M., Botstein, D., 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42, 385–413.

Tse-Dinh, Y.C., Beran-Steed, R.K., 1988. *Escherichia coli* DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains. J. Biol. Chem. 263, 15857–15859.

Van Mellaert, L., Mei, L., Lammertyn, E., Schacht, S., Anné, J., 1998. Site-specific integration of bacteriophage VWB genome into *Streptomyces venezuelae* and construction of a VWB-based integrative vector. Microbiology 144, 3351–3358.

Wagner, P.L., Waldor, M.K., 2002. Bacteriophage control of bacterial virulence. Infect. Immun. 70, 3985–3993.

Walderich, B., Holtje, J.V., 1991. Subcellular distribution of the soluble lytic transglycosylase in *Escherichia coli*. J. Bacteriol. 173, 5668–5676.